Assimilation of multiple datasets results in large differences in regional to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model

C. Bacour, N. MacBean, P. Peylin, F. Chevallier

with contributions from S. Léonard, E. Koffi

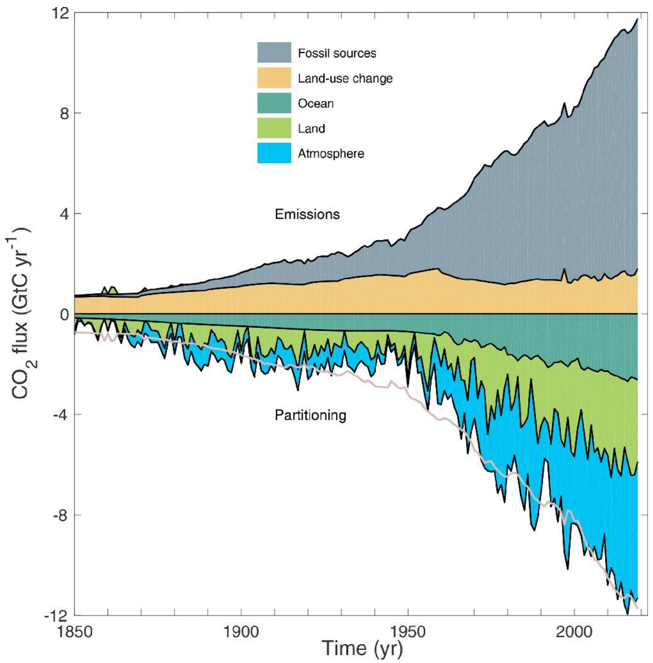
Laboratoire des Sciences du Climat et de l'Environnement
Gif sur Yvette, France

in Biogeosciences - Discussion
https://doi.org/10.5194/bg-2022-109

Terrestrial vegetation: the most uncertain component of the Global Carbon Cycle

Quantifying and reducing uncertainty in global C budget projections using Terrestrial Biosphere Models

The Global Carbon Budget



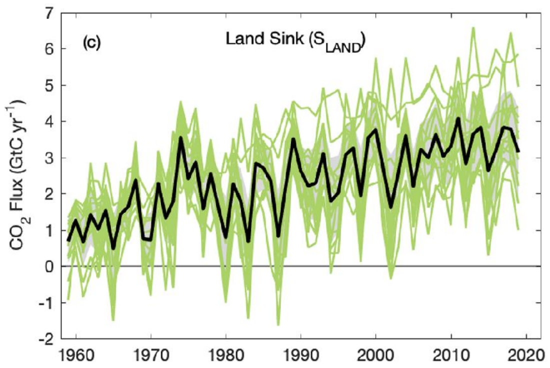
@Friendlingstein et al. (2020)

3.4 ± 0.9 $GtC.yr^{-1}$
over 2010-2020

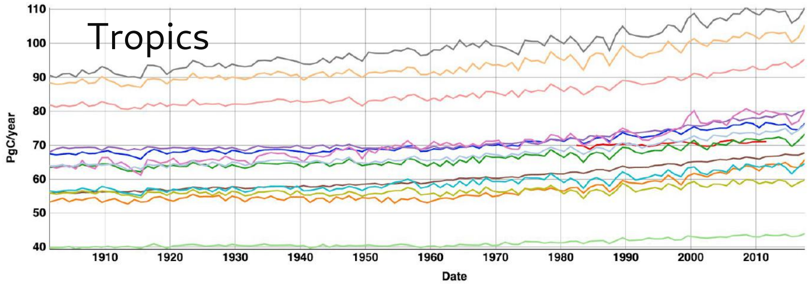
NEE = GPP-TER

Simulations by Terrestrial Biosphere Models

land C sink - NEE



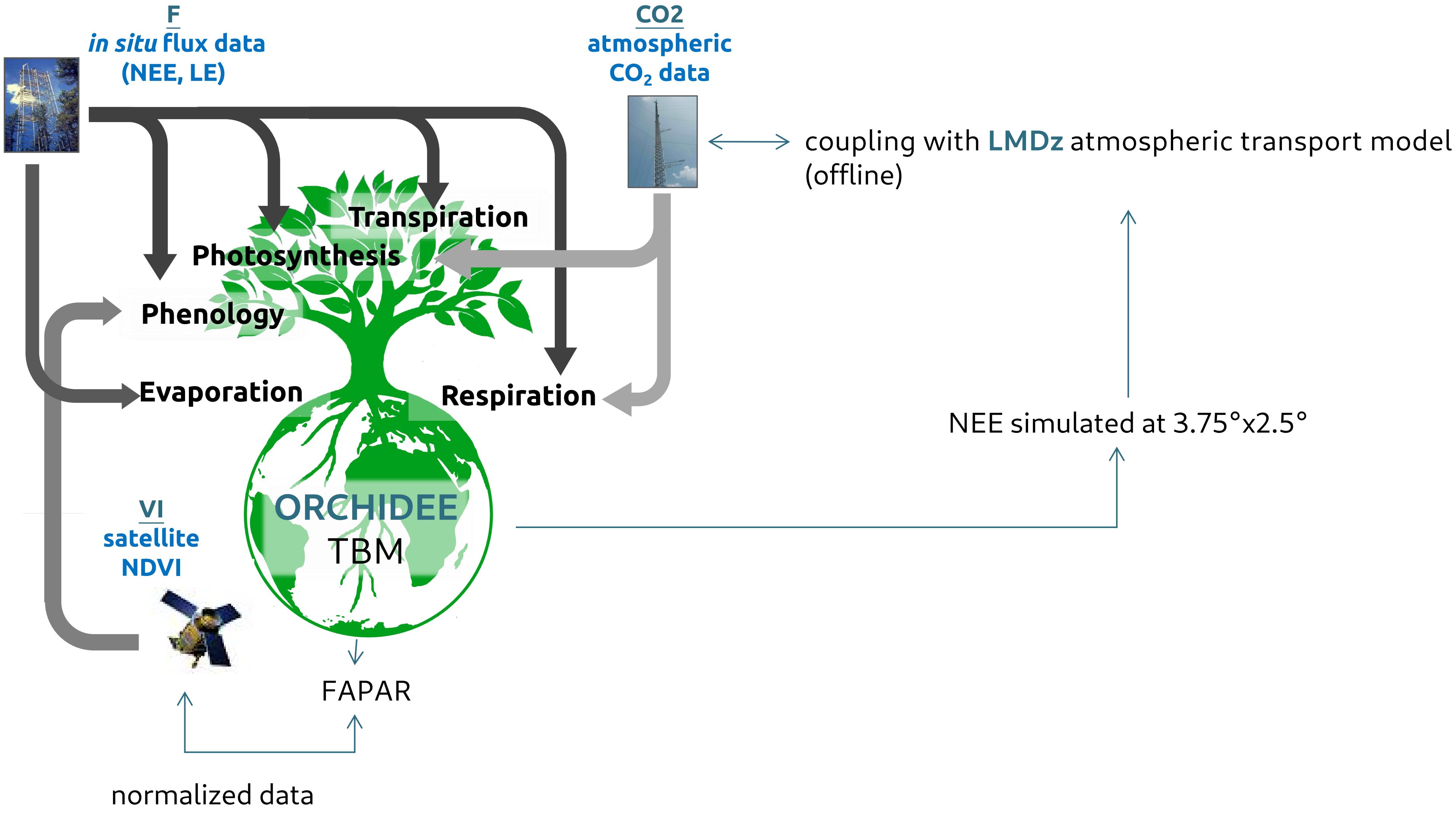
Gross C flux (GPP)



results from TRENDY V9

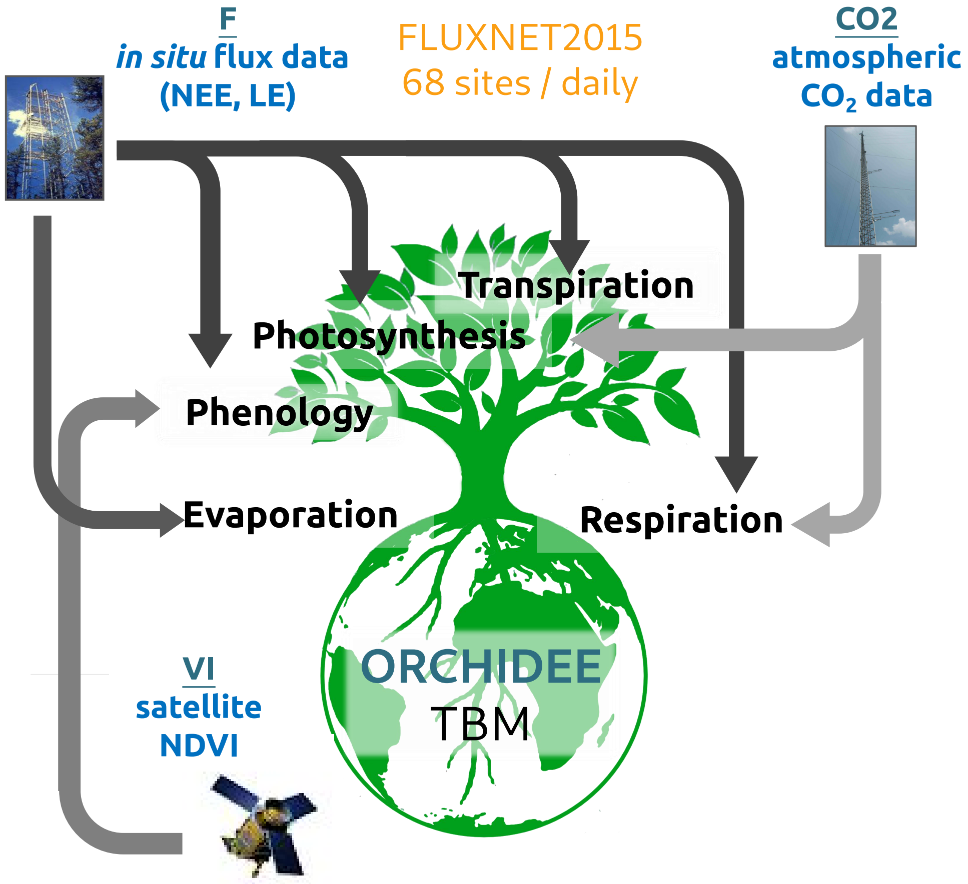
Model - data fusion to optimize ORCHIDEE parameters

Datasets



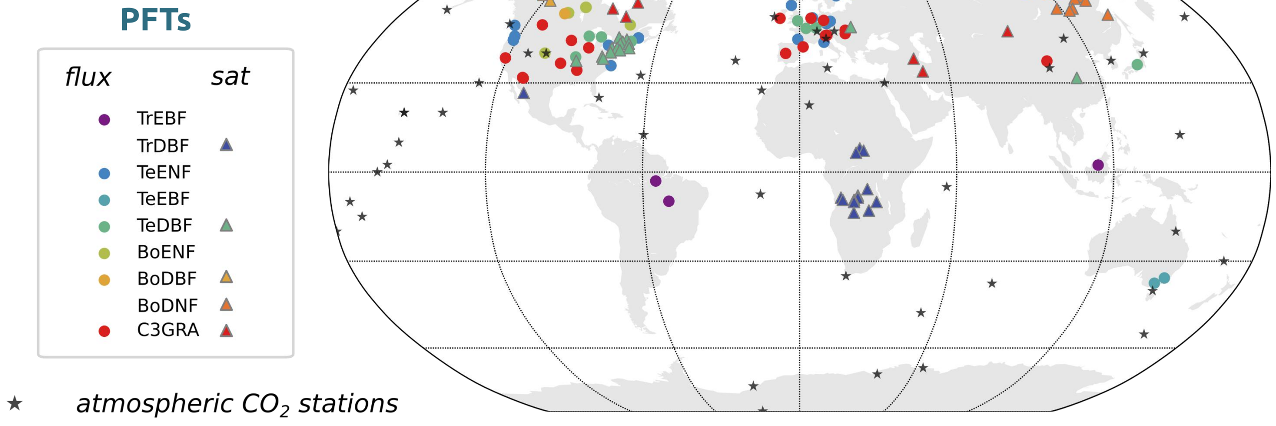
Model - data fusion to optimize ORCHIDEE parameters

Datasets



MODIS coll. 5
15 pixels / PFT
daily / 0.72°
2000-2010

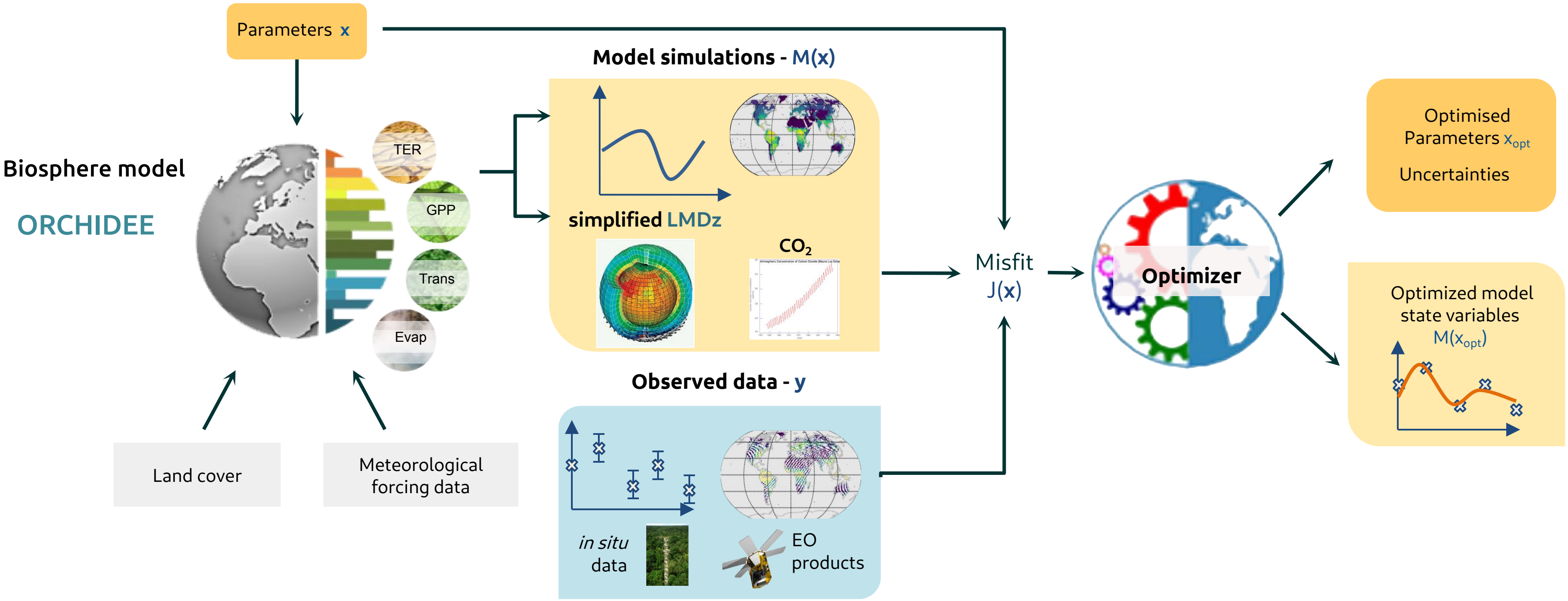
53 surface stations
monthly
2000-2009



- Not an up-to-date reanalysis of the C cycle!
- Assessment of the useful informational content provided by different data-streams on global C fluxes wrt the set-up and the model structure

ORCHIDAS assimilation framework

https://orchidas.lsce.ipsl.fr



focus on NEE & GPP

Misfit function

$$
\begin{aligned}
J(\mathbf{x}) \;=\; & \frac{1}{2}\Big[\left(H_{LMDz}{\circ}H_{ORCH}(\mathbf{x})-\mathbf{y}^{CO2}\right)^{t}.\mathbf{R}^{-1}_{\mathbf{CO2}}.\left(H_{LMDz}{\circ}H_{ORCH}(\mathbf{x})-\mathbf{y}^{CO2}\right)+ \\
& \left(H_{ORCH}(\mathbf{x})-\mathbf{y}^{F}\right)^{t}.\mathbf{R}^{-1}_{\mathbf{F}}.\left(H_{ORCH}(\mathbf{x})-\mathbf{y}^{F}\right)+ \\
& \left(H_{ORCH}(\mathbf{x})-\mathbf{y}^{VI}\right)^{t}.\mathbf{R}^{-1}_{\mathbf{VI}}.\left(H_{ORCH}(\mathbf{x})-\mathbf{y}^{VI}\right)+ \\
& \left(\mathbf{x}-\mathbf{x}^{b}\right)^{t}.\mathbf{B}^{-1}.\left(\mathbf{x}-\mathbf{x}^{b}\right)\Big]
\end{aligned}
$$

Data Assimilation Experiments

Experiment name	Flux data	NDVI data	Atmospheric CO_2 concentrations	Number of optimized parameters	Number of observations
F	x			133	150 792
VI		x		19	149 916
CO2			x	114	6 360
F+VI	x	x		152	300 708
F+CO2	x		x	182	157 152
VI+CO2		x	x	114	156 276
F+VI+CO2 **F+VI+CO2-2steps**	x	x	x	182	307 068

Optimized parameters

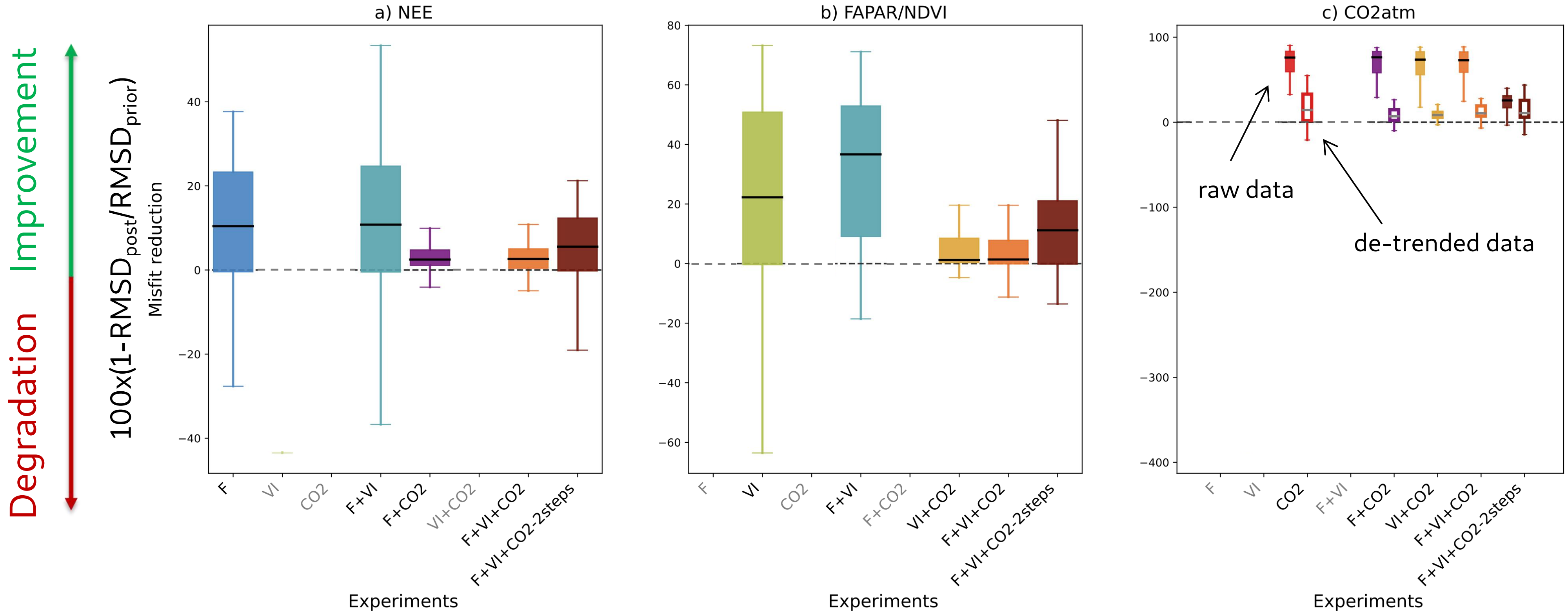
Processes	Parameters	Obs. constraint
Photosynthesis	5 parameters / PFT	F,CO2
Soil Water Availability	1 parameter / PFT	F,CO2
Phenology	5 parameters / PFT	F,VI,CO2
Respiration	3 parameters + $KsoilC_{site}$ + $KsoilC_{reg}$ (30 regions)	F,CO2 F CO2

Scientific Questions

- Analyze the compatibility, complementarity, and usefulness, of the different data streams in the frame of a global-scale C data assimilation system
- Assess their potentials to improve the realism of the space-time distribution of **NEE** and **GPP**

Overall fit to the observations

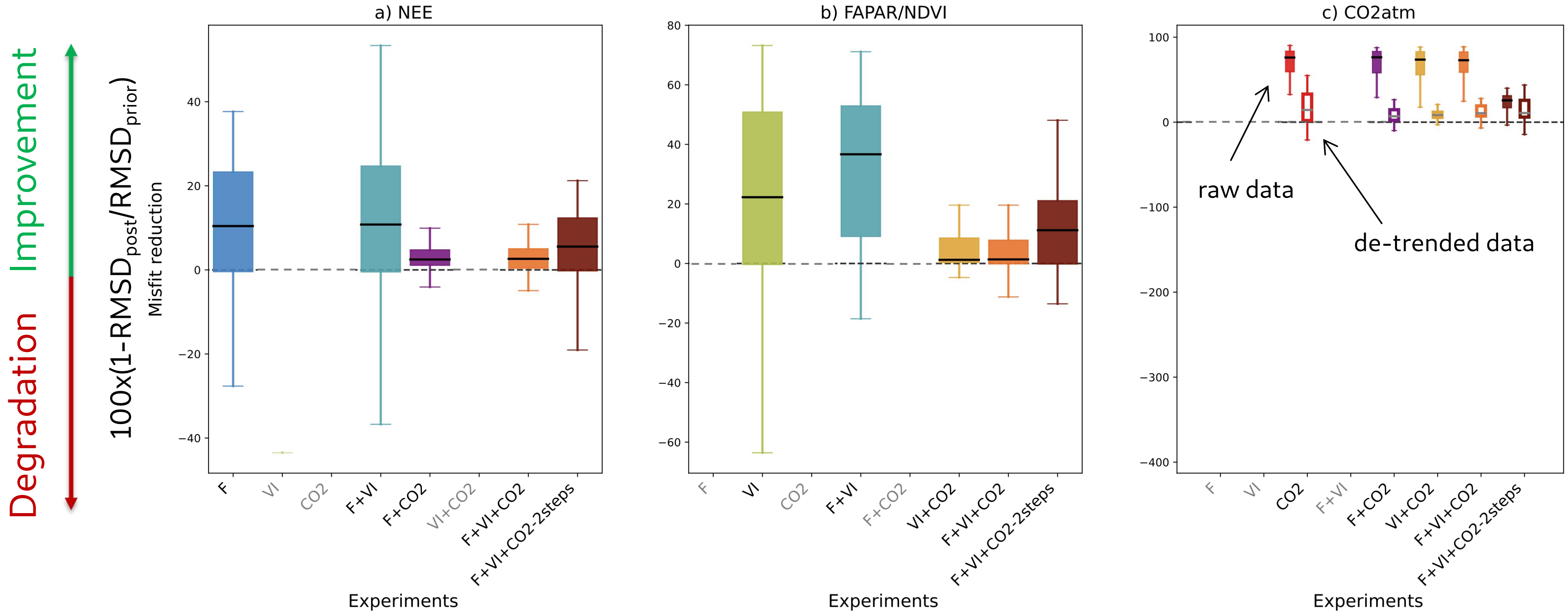
Reduction of the model-data mismatch wrt the variables assimilated



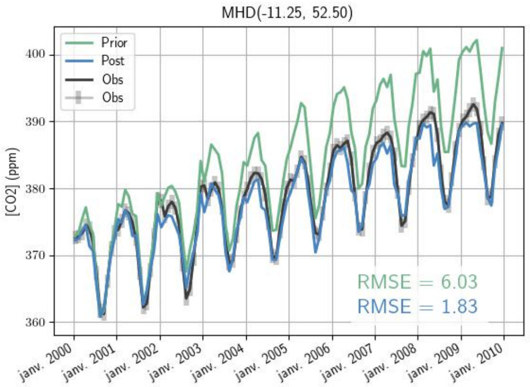
- Model improvement usually better for the experiments where a data stream is assimilated alone
- For the multiple DA with CO2, the 2-step approach lead to the highest improvement wrt F and VI
- For CO2: slight differences wrt raw data, higher variability between experiments wrt de-detrended data

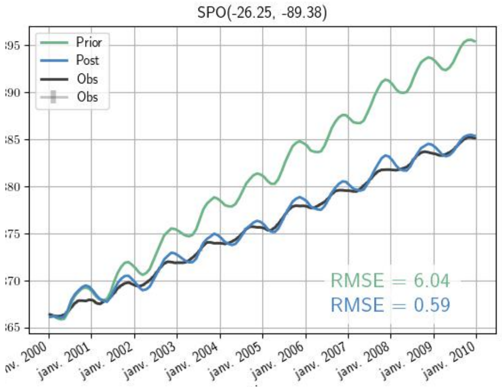
Overall fit to the observations

Reduction of the model-data mismatch wrt the variables assimilated



atmospheric CO_2 time series

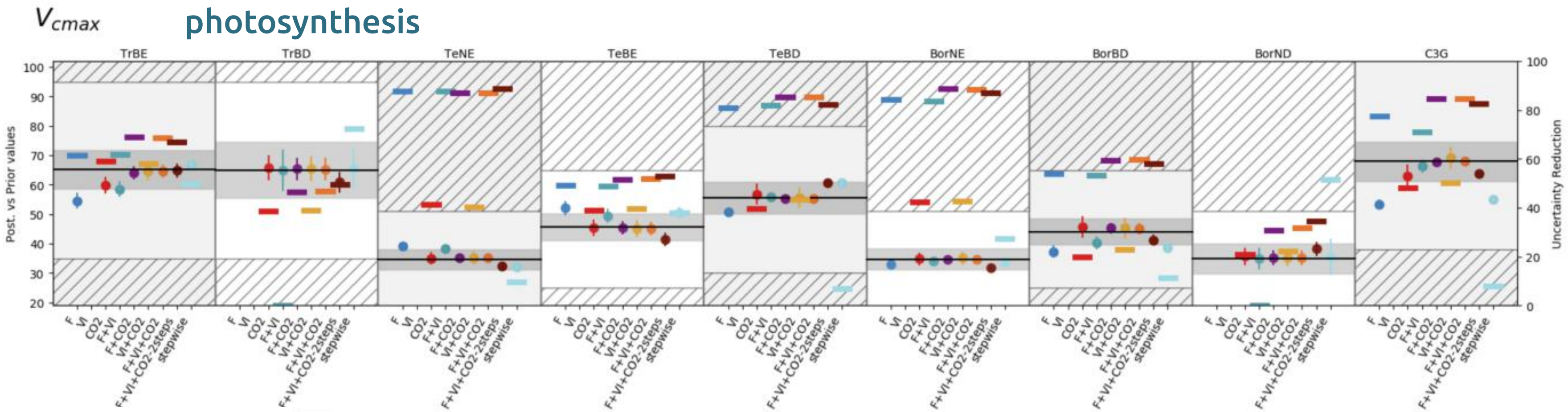




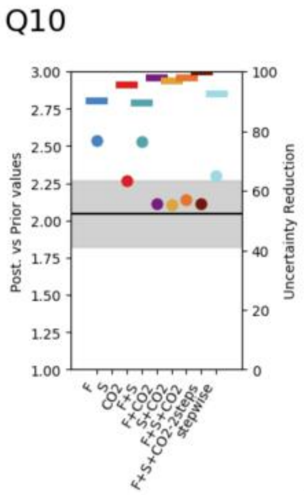
trend correction mostly achieved by the correction of the soil C disequilibrium (multiplicative $KsoilC_{reg}$ parameter)

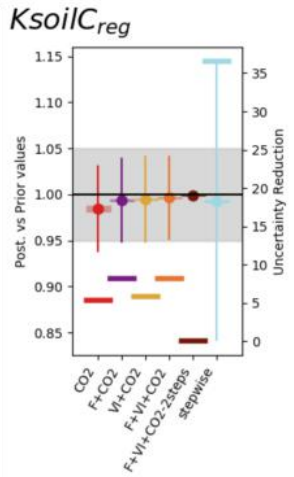
Parameter estimates and uncertainties

photosynthesis



respiration





- Highest departures from the prior values obtained for single-data stream assimilations
- Correcting the bias in atm. CO_2 trend prevails over the improvement of photosynthesis and phenology related parameters
  - higher changes obtained for F and VI, compared to CO_2
  - little variability among the 3 data-streams DA experiments (**but the 2steps one**)

Relative constraints brought by the different datasets

Influence matrix

$$\mathbf{S} = \mathbf{R}^{-1}\mathbf{H}^{\infty}\mathbf{A}\mathbf{H}^{\infty\mathbf{t}} \qquad \text{with} \qquad \mathbf{A} = [\mathbf{H}^{\infty\mathbf{t}}\mathbf{R}^{-1}\mathbf{H}^{\infty} + \mathbf{B}^{-1}]^{-\mathbf{1}}$$

(@ *Cardinali et al., 2004*)

Global Observation Influence (OI)

> gauges the average influence that each single observation has on the analysis

$$OI = \frac{tr(\mathbf{S})}{m}$$

Relative Degrees of Freedom for Signal (DFS)

> measures the relative contribution of the data stream *o* to the fit

$$DFS = 100 \times \frac{tr(\mathbf{S})}{tr(\mathbf{S_o})}$$

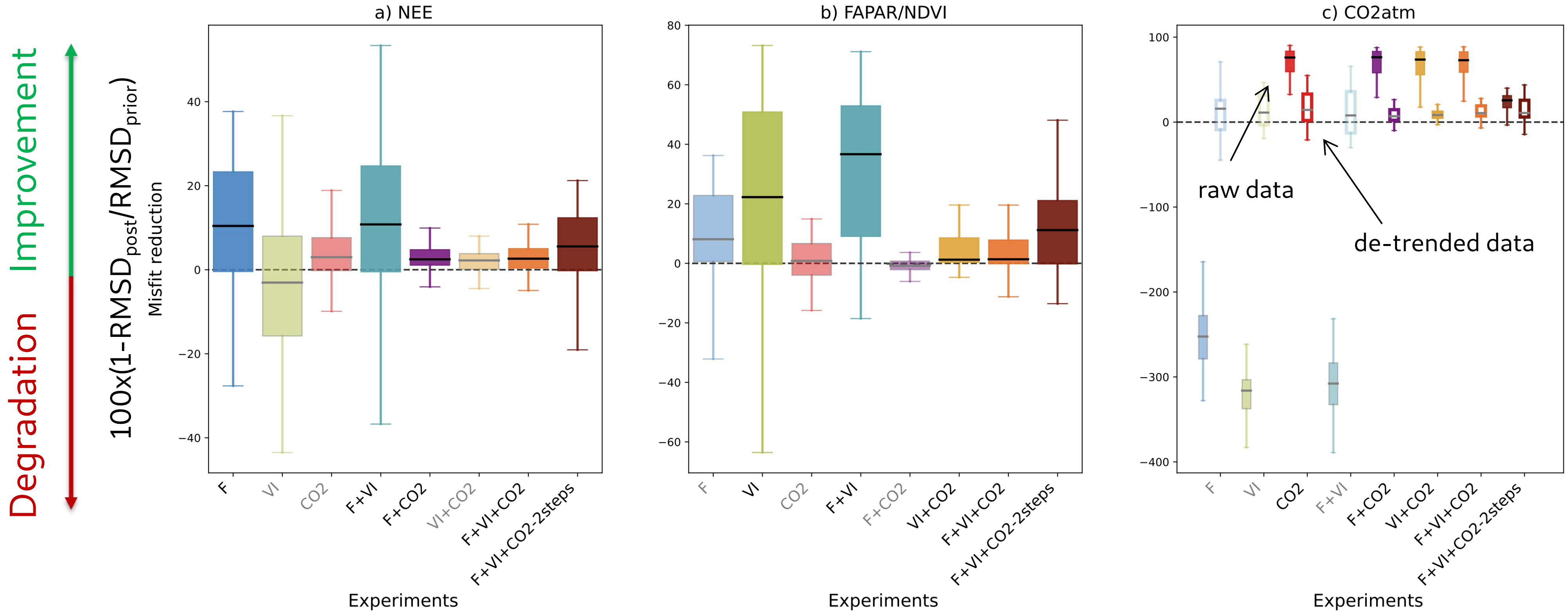
	OI		DFS	
	1-step	2-step	1-step	2-step
F	0.000586	0.000577	74.65	76.9
VI	0.000048	0.000048	11.12	11.68
CO2	0.002654	0.002035	14.23	11.42

~ x 5

- OI for CO_2 is about 5 times higher than that for F with about the same number of optimized parameters BUT about 25 times less observations > impact of the trend correction

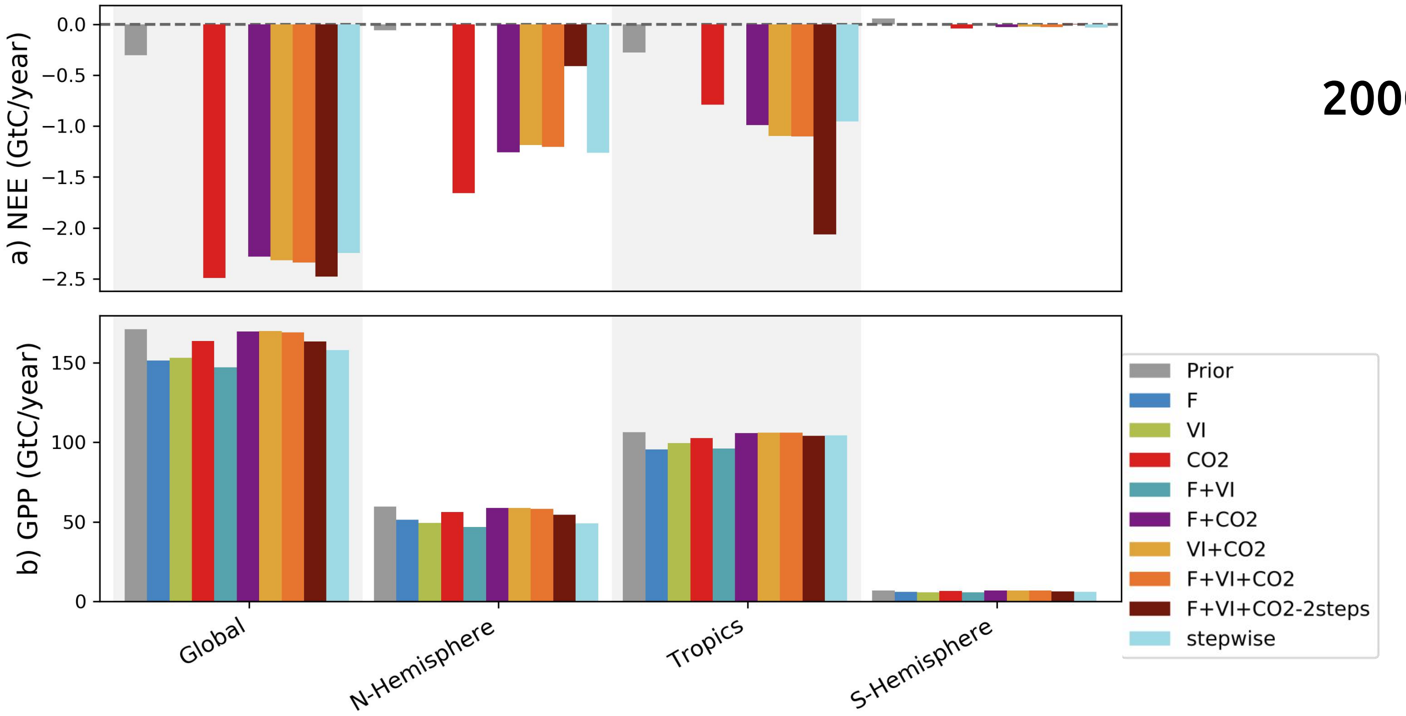
Overall fit to the observations

Reduction of the model-data mismatch for all variables



- Simulations using optimized parameters constrained by other data-streams > model degradation can be observed
- The joint assimilation experiments lead to improved model-data agreement and reduce the risk of model degradation (model overfitting wrt a given data-stream)

Impact of the assimilations on regional to global land C fluxes



2000-2009

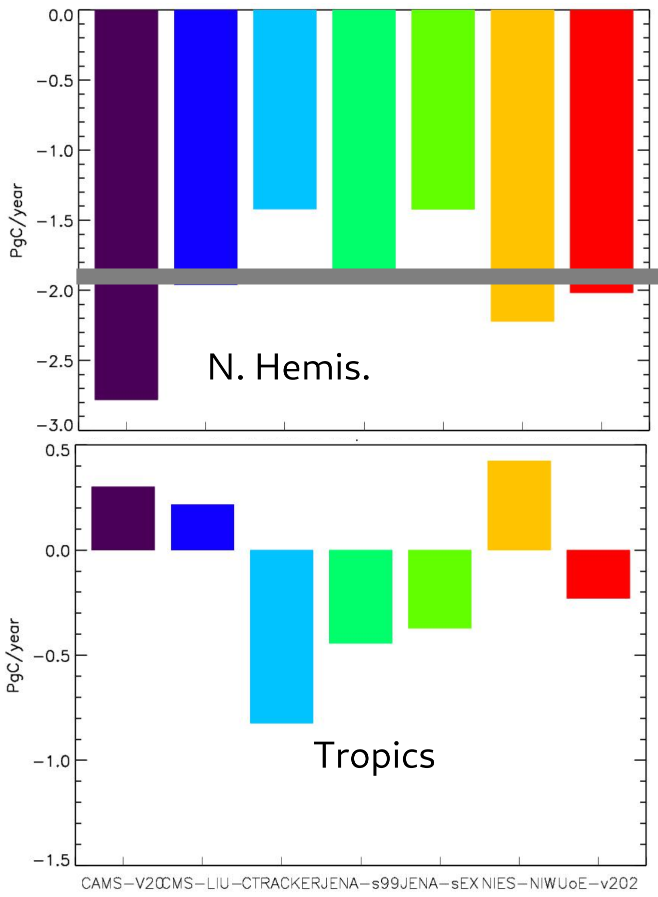
GCP: NEE = -2.9 GtC.yr^{-1} ± 0.8

- **F+CO2**, **VI+CO2**, and **F+VI+CO2** : similar NEE and GPP budgets across regions
- **CO2** and **F+VI+CO2-2step** experiments result in distinctly different estimates between N. Hemisphere and Tropics

Different NH - Tropics partitioning estimated by atmospheric inversions vs TBMs

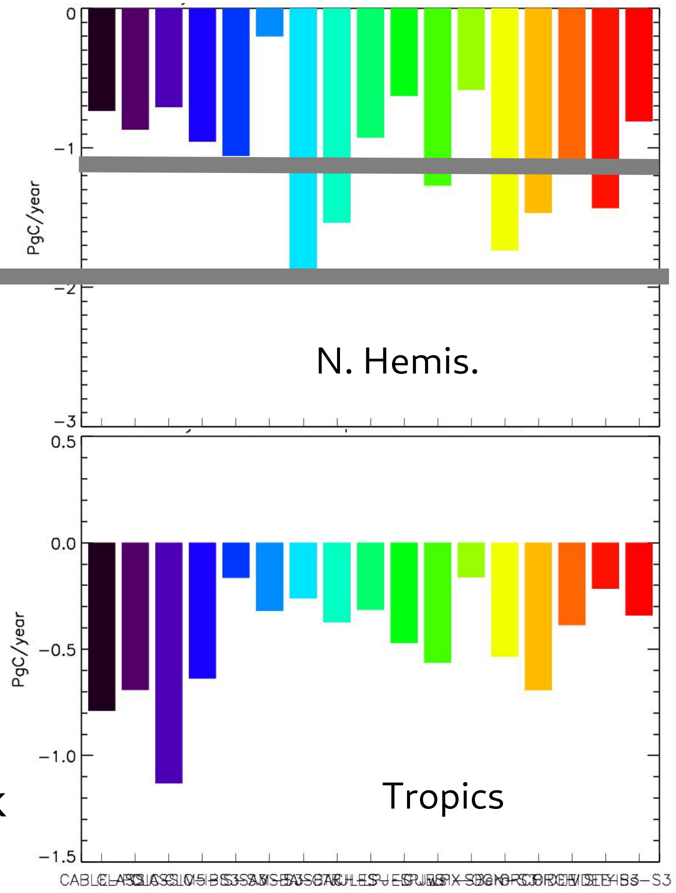
net C fluxes:
Mean over
2010-2020

Atm. inversions (GCP-2021)



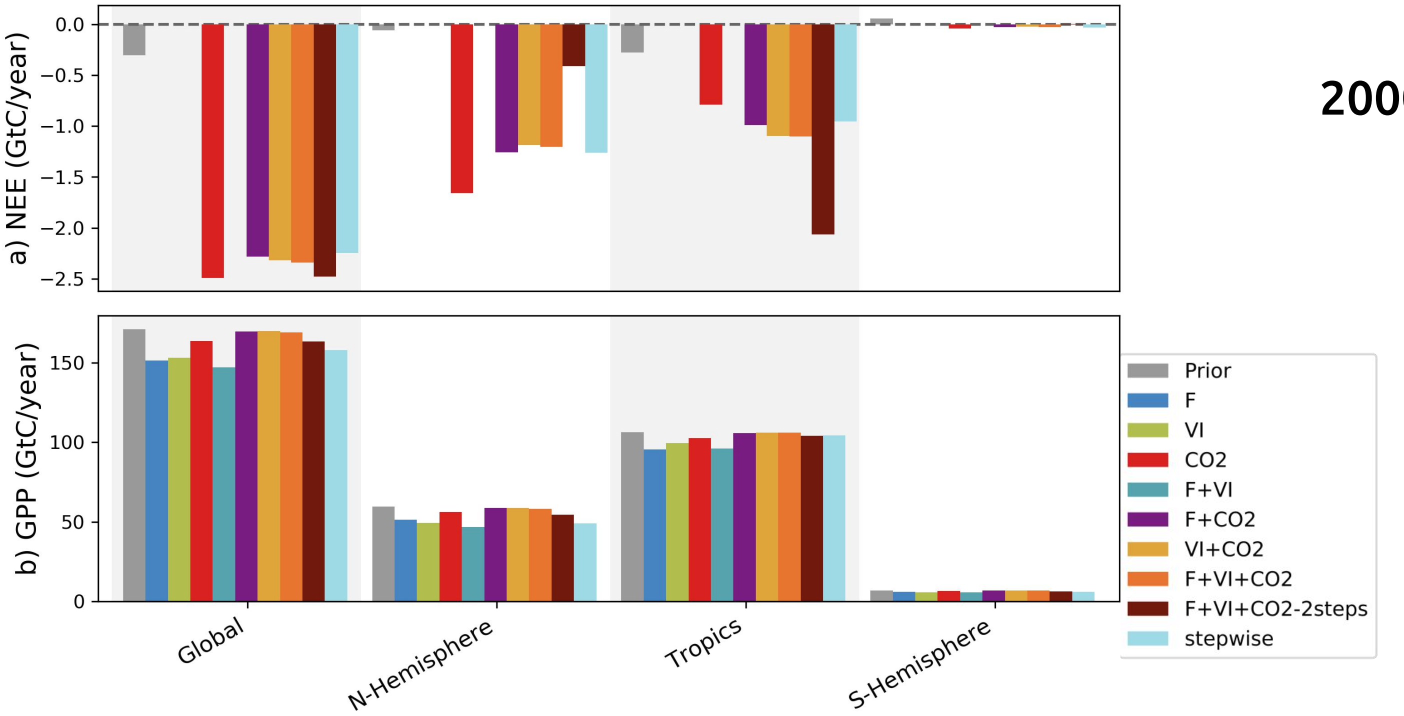
larger sink

TBMs (TRENDY - V10)



larger sink

Impact of the assimilations on regional to global land C fluxes



2000-2009

GCP: NEE = -2.9 $GtC.yr^{-1}$ ± 0.8

C budget partitioning between N. Hemisphere and Tropics

- **CO2** > similar regional partitioning as the atmospheric inversions
- **F+VI+CO2-2step** > typical partitioning pattern of TBMs' behavior
- **F+CO2**, **VI+CO2**, and **F+VI+CO2** > approximately equal C sink in the NH and tropics (=> unlike the general pattern for TBMs)

Take home messages

- Configuration matters

- Atmospheric CO_2 data are crucial for an accurate prediction of the distribution of the terrestrial land sink
  - challenges in handling model-data bias in Bayesian optimisation frameworks
  - sub-optimal optimization of the soil C disequilibrium with our approach based on a model spin-up without a long transient run (not TRENDY like)
  - 2step approach: illustrate how the informational content of the data-streams relative to C fluxes is enhanced once soil C disequilibrium is modeled in a more "realistic" way

- Diagnostics for system evaluation
  - relative informational content brought by each data stream
  - consistency of the error statistics on parameters and observations (Desrozier et al. (2005))
  - optimisation efficiency

- Assimilating simultaneously multiple datasets is preferable to optimize the values of the model parameters and avoid model overfitting